This workshop introduces the development of model estimation data sets, the structures of the various model components, and the procedures for estimating models. At the end of the series, participants will be able to better manage model development done for them by others (e.g., consultants) and to understand and evaluate the results.

Questions and Answers during Session IV

Q: How would you determine the peak period, based on departure time, or arrival time, or the actual trips on the roads? How would you handle trips longer than 1 hour?
A: This is a great question...we skipped this part for want of time...typically, we define time of day based on arrival time or departure time, but one could also use the mid point of the trip to classify the trip into a given bin. This should take care of the trips which are longer than one hour.

Q: Thanks for your response. I note that this is being recorded and I am hoping that this session will be able on the net sometime soon?
A: Absolutely! The recording should be on the web soon along with the slides, homework and homework solutions.

Q: Can logit model be used for multi modal trips such as SOV and mass transit?
A: I am assuming you are referring to "drive access to transit trips"...if this is the case, then one can define this as a separate choice and estimate models accordingly...in fact, this is done quite often in many urban areas.

Q: what would be my purpose if I drop my kid to school and go further to my work place? Or first I go to a shop before I go to my work?
A: Excellent question. This is subject to considerable debate, and in fact points to a problem with the traditional definitions of trip purposes. The way to label your examples within the context of a four-step model is to call them "journey to-work" trips because both the legs of the trip have been pursued en route to work.

Q: Is the utility function a linear regression equation with error term e and regression coefficients b?
A: The utility expression for each mode is a linear combination of various attributes that include mode-specific attributes, person attributes...however, note that this is not a linear regression...this is a logistic regression...the difference will be apparent to you in a moment

Q: Could you shed some light on acceptable magnitudes of the alternative specific constants?
A: Another very good question. The alternative specific constant can be thought of as a catch-all for everything that you were NOT able to capture using your other variables....there are no hard and fast rules about the ASCs...but clearly, they should not be too highly negative or too highly positive...because then they would overwhelm the effect of the other explanatory variables...the FTA has some criteria about how big the rail constants can be as compared to bus etc...but again, these are subjective.

Q: if you are estimating a route choice model that partitioned auto trips, would you expect a higher value of time than if you were estimating a mode choice model with a broader spectrum travelers (such as low income, transit dependent observations)?
A: Absolutely! the VOTs do change based on the income of the individual...for example, high income travelers may actually have a higher implied VOT than do the lower income travelers...it

is indeed good practice to segment mode choice coefficients such as time and cost based on the income of the individual...we will see this in the homework today...

Q: Is it better to have toll and non-toll nests in mode choice, or post-process toll and non-toll paths in traffic assignment, or both?
A: We have seen this done both ways...unfortunately, there is no clear answer or consensus. If one can estimate a nested model successfully with toll and non-toll modes and has the data to do so, then it may make sense to make this distinction within mode choice.

Q: Is it necessary that the two utility functions for alternatives be the same? Or can each utility function for each alternative have different number of variables?
A: It is definitely not necessary for the utilities of two alternatives to have the same variables.

Q: Is Toll route choice model an example of Logit models? What's your preference about choosing toll choice first or mode choice first?
A: Please see my response to the earlier question...route choice models can certainly be modeled as discrete choice models...but, the simple MNL which we are discussing here is not adequate to model route choice...

Q: Now I see Yasasvi! Coding for deterministic variables matters in regression but not in logit, probit, etc. Now I see!!
A: You bring up a good question...we will address this when we open up for questions in a moment.

Q: How significant is the IIA property in practice? Is there a defensible reason for using a MNL model even if two alternatives are similar?
A: If two alternatives are similar...it is not a good idea to use an MNL...people do it anyway because an MNL is the easiest to estimate!

Q: I do not know if the session would discuss on marginal effects? Please see if it is possible to discuss on this for couple of minutes.
A: Do you mean the price elasticities of mode choice?

Q: Not exactly price but effect of percentage changes in some of the variables.
A: Yep...exactly...the percentage change in say the share of a mode due to a percentage change in price or time...this is an Excellent point...I think Tom will bring this up at some point...if not, please feel free to raise this at the end of the presentation.

Q: Is the Logsum a way to obtain MLE?
A: Yes. This is certainly one way to look at the logsum...ultimately, when we formulate the log-likelihood and maximize it to obtain the parameters, we use the logsum in this calculation...so you are right.

Q: Could you please repeat the logsum explanation?
A: Please let us know if this is still not clear...we can talk about this at the end.

Q: Could you discuss a little bit about how multinomial logit model and nested logit model work differently in a mode choice model?
A: The mode choice model is most appropriate when your modes are very dissimilar...say, auto, transit and walk...the nested logit models are used when you have some modes that are more like each other than the rest...so for instance, let's say you want to model, auto, bus, rail with

drive access and rail with walk access...you know for sure that the two rail modes are very similar to each other...this is why you define them as siblings within a nest.

Q: Skims used to produce level of service imply a trip table, which already implies a logit mode split model. Isn't there a lack of consistency here?
A: I think you are referring to the generation of skims used for mode choice...this is a very good question...the ideal way to resolve this problem is as follows: Step 1: Use free flow skims for each mode to estimate a mode choice model Step 2: use the mode choice model to estimate demand by each mode (that is the trip table for each mode) and then re-estimate the new skims...Step 3: Go thru a second iteration of estimation...

Q: I am wondering whether logsum came from a natural 'log' of 'sum' of exp(v). Why is V used instead of U for utility?
A: We will discuss this at the end...I think this might be better explained on the phone

Q: Should you run correlations first (Spearman's rho) to look at multicollinearity before adding variables to the logit model?
A: Excellent point...Typically most modeling software gives you a correlation matrix for all the variables used in the model...depending on what you observe here...you could choose to drop some variables that seem highly correlated to the others in the specification.

Q: At the end of the lecture please discuss (1) normality of variable data (2) affects of missing data (3) the use and interpretation of residuals.
A: Time permitting...we will certainly do so.

Q: Is there a base formula for determining VOT?
A: yes...it is (coefficient of IVTT)*60/Coefficient of Cost ....where the cost is expressed in dollars and time in minutes.

Q: This is nitpicking a bit, but you can not assume that ALL people would always value a 5 minute car ride over a 15 minute walk, for example. The 15 minute walk may take longer, but some people will see it as exercise. The same could be true of a bike ride vs. a much shorter car ride. These assumptions always come up in value-of-time discussions. Some people prefer a 1 hr bus ride (or train ride) over a 30 min car ride, because they can sleep/talk on the phone/read on the bus or train.
A: Excellent points...the problem with econometric models is that there are always unobserved characteristics that an analyst can never observe...so, for example, the preference for exercise is not something that can be observed or quantified...this is why the best calibrated models can only have a rho-squared of 25%!!! In short, the VOT we get from a model is an average value across a whole bunch of people...it must be interpreted in the context of things you rightly pointed out...

Q: What is the unit for transit/highway accessibility ? (income, 1000 dollars,)
A: This variable is the ratio of the percentage of total regional employment which can be reached in 80 minutes by transit from the origin zone to the percentage of total regional employment that can be reached in 60 minutes by highway from the origin zone.

Q: are the coefficients in that table more like b's or betas?
A: yes...these are the estimated regression coefficients...

Q: how is rho squared w.r.t. zero determined?
A: Good question...one can think of the zero-coefficients model as the most naive model...so the rho-squared w.r.t to zero tells us how we are performing w.r.t the most naive model...it is defined as the (Likelihood of the model - likelihood of zero coeff model)/(likelihood of zero coeff model).

Q: With respect to Tom's Vehicle Availability Model, it looks like "Persons less than vehicles" is derived from other variables already included in the model. Is that not in a sense double counting those variables?
A: You are right...some of the variables are related...but let's consider persons/household and persons less than vehicles...these two variables are seemingly related...but they capture slightly different effects.

Q: Going back to my previous point, the constant for transit walk access for example, in terms of IVT is about 340 minutes. How does one justify a constant this large?
A: The cost here is in dollars and the time is in minutes...so the formula is -.02*60/-.4 = 3 dollars/hour.

Q: if you have a nested model such as this and the two nests are not significant, should you roll up the model to just have transit as a choice rather than dividing it into walk and auto access?
A: absolutely!

Q: Can you have coefficients for the same variable vary by modes?
A: Yes, this is definitely possible.

Q: What is the unit for transit/highway accessibility?
A: This variable is the ratio of the percentage of total regional employment which can be reached in 80 minutes by transit from the origin zone to the percentage of total regional employment that can be reached in 60 minutes by highway from the origin zone.

Q: web biogeme: http://transp-or.epfl.ch/page63023.html

Q: Will an IVT coefficient of -0.02 and a Constant of 6.83 (yielding a bias value of 341.5 minutes) raise eyebrows?
A: if this coefficient is for bus...then it might...but the coefficient needs to be evaluated in the context of the other coefficients.

Q: How can we account for some external factors like weather, e.g. during harsh winter/snow conditions, walking/bicycle as a mode has very diminished utility?
A: Did we answer your question?

Q: You might have misunderstood my question. I was referring to slide 35, constant/IVT. I wasn't referring to the VOT.
A: My bad...you are right! This is indeed a case where the constant implies a very high value...but notice that there are lots of variables in the utility expression that have negative coeffs.

Q: Yes, thanks. The reason for asking the question is that winter in some regions lasts 25%-33% of the year.

A: Tell me about it! The issue is that essentially you would have two models, winter and non-winter, and that means twice as much data, twice as much application, etc. So far, agencies, even up here in New England, have not decided to allocate their resources that way.

Q: What are the software options for estimating these models?
A: Did we answer this sufficiently?

Q: where can I find the home work info?
A: You will find this at the website...you will also receive an email from Gary Thomas once all the material is posted...

Q: What is the best way to estimate value of time in dollars?
A: coefficient of IVTT * 60/coefficient of cost (the 60 assumes the IVTT is expressed in minutes – this expression therefore gives value of time in $/hour).

Q In what situations are cross-nested logit models recommended?
A: Cross nested models have not yet made it into standard practice but have great potential. Basically, they are useful when some alternatives are similar to more than one other alternative, such as park and ride being similar to both auto and transit-walk access.

Q: I would like to know about marginal effects? The percentage change in probability of choice with change in a percent change in some of the variables. I also remember reading somewhere that this is done differently for continuous variables and for non-continuous variables.
A: We haven't had the time to answer your question. Our apologies...can you please write to ypopuri@camsys.com or trossi@camsys.com, we will respond to your email in detail.